Chemistry Education Research and Practice

PAPER

View Article Online
View Journal | View Issue

Cite this: *Chem. Educ. Res. Pract.*, 2018, **19**, 885

Student-centred active learning approaches to teaching quantum chemistry and spectroscopy: quantitative results from a two-year action research study

Lauri Partanen [a,b]

In this article, I propose a student-centred approach to teaching quantum chemistry and spectroscopy at the bachelor-level that extends active learning principles outside course lectures. The aim is to elucidate what type of methodology is most appropriate and efficient for this context and student population, and how this incorporation of active learning elements impacts learning. Three quantitative learning indicators are used to measure the effectiveness of the proposed approach, including exercise points obtained by the students, exam results, and the results of a conceptual inventory administered both at the beginning and the end of the course. The proposed model resulted in substantial improvement in learning outcomes compared to a previous class where active learning elements were confined mostly to the course lectures and a traditionally taught class. The model can be generalised to any subject where both quantitative and qualitative understanding is required. Thus, in addition to providing further support for the effectiveness of active learning approaches in science, this study shows the benefits of applying these approaches to exercises and other course tasks besides lectures.

Received 12th March 2018,
Accepted 28th May 2018

DOI: 10.1039/c8rp00074c

rsc.li/cerp

1 Introduction

Quantum mechanics persists as one of the most challenging topics many chemistry and physics students face during their bachelors' degree studies. These challenges arise partly from the large number of unfamiliar and unintuitive concepts and the profusion of mathematical expressions that characterise it (Tsaparlis, 2001; Singh 2008). The ubiquitousness of equations makes quantum mechanics especially challenging for chemistry students whose mathematical foundations are often poorer than their physics counterparts'. Regardless, basic knowledge of quantum mechanics is vital for understanding virtually all chemical processes as well as the functioning of lasers and many other instruments that chemists and physicists utilise in their daily practices.

Multiple approaches to teaching quantum mechanics have been developed to help students come to grips with this difficult subject. For example, Tsaparlis together with Sánchez-Gómez and Martín have proposed using a historical perspective, where the postulates of quantum mechanics are made more acceptable by connecting them to their origins, thus making visible the number of alternative models that were available at their conception (Tsaparlis, 2001; Sánches Gómez and Martín, 2003). Other teachers have suggested the adoption of a Bohmian framework to provide a continuation with earlier ideas and help students create links between new and existing knowledge (Passon, 2004). This is because Bohmian mechanics preserves the concept of a particle trajectory, while the wave function acts as a guiding field that gives rise to quantum mechanical interference phenomena.

Several authors have argued for a qualitative focus in the teaching of quantum mechanics at the introductory-level (Müller and Wiesner, 2002; deSouza and Iyengar, 2013). For example, Kalkanis *et al.* proposed placing students' initial alternative conceptions to the centre of the instruction by juxtaposing the starkly different worldviews of classical and quantum mechanics to facilitate the radical conceptual change necessary (Kalkanis *et al.*, 2003). Despite the differences in these worldviews, Marshman and Singh posit in their framework for understanding patterns of student difficulties in quantum mechanics that the challenges many students face in developing expertise in these two fields are analogous (Marshman and Singh, 2015): They both require similar overhauls in reasoning and problem-solving capabilities. In many of the qualitative approaches, student interaction with simulations and virtual laboratory experiments allows for the first-hand exploration of quantum mechanical phenomena and is a key component in conceptual

[a] *Department of Chemistry, University of Helsinki, P.O. Box 55 (A.I. Virtasen aukio 1), FI-00014, Finland*
[b] *Department of Chemistry and Materials Science, Aalto University, P.O. Box 16100, FI-00076 Aalto, Finland. E-mail: lauri.partanen@helsinki.fi*

development (Müller and Wiesner, 2002; Kalkanis *et al.*, 2003; Kohnle *et al.*, 2010). Indeed, both in introductory physics and quantum mechanics, computer animations and simulations have been shown to be effective learning tools, particularly supporting the development of mental representations of physics concepts (Finkelstein *et al.*, 2005; McKagan *et al.*, 2008; Adams *et al.*, 2008a, 2008b; Podolefsky *et al.*, 2010). Besides simulations, laboratory experiments can also illuminate critical concepts and support qualitative understanding in quantum mechanics, although they typically require much more resources (Galvez *et al.*, 2005).

More generally, active learning and student centred-approaches to teaching are becoming commonplace in science due to the rapidly growing literature attesting to their effectiveness (Prince, 2004; Slunt and Giancarlo, 2004; Ruiz-Primo *et al.*, 2011; Wright, 2011; Freeman *et al.*, 2014). Prince (Prince, 2004) distinguished four different types of active learning: active learning practices adopted in lectures (or simply active learning in many contexts such as Freeman *et al.* (Freeman *et al.*, 2014)), collaborative learning, cooperative learning, and problem-based learning. He defined active learning as any approach that engages students in the learning process and requires students to think and participate in meaningful learning activities. In contrast, student-centred approaches shift the focus from the teacher to the learner by readjusting the balance of power in the classroom and the roles of the teacher and the student, reconsidering the function of course content, and re-evaluating student responsibility for learning together with the purpose and processes of evaluation (Weimer, 2002; Wright, 2011).

In this article, I offer a detailed account of a student-centred and research-based course structure for teaching quantum mechanics and spectroscopy, developed as a part of an action research (Ralle and Eilks, 2002; Tripp, 2005; Gibbs *et al.*, 2017) initiative on a second-year bachelor-level course in chemistry at the University of Helsinki. The changes to the course were undertaken in two stages: in 2016 the focus was on incorporating active learning practices to the course lectures and in 2017 to the exercise structure. For comparison, data from the traditionally taught 2015 course is also presented. The impact of these changes was evaluated by employing a mixed-method approach (Johnson *et al.*, 2007), with both quantitative and qualitative indicators. Assessment and development of the student's learning objectives in different parts of the course structure were done in the framework of the revised Bloom's taxonomy (Bloom, 1956; Anderson and Krathwohl, 2001). The taxonomy is also employed to compare and contrast different course tasks based on the chemistry specific classification suggested by Tikkanen and Aksela (Tikkanen and Aksela, 2012) (See Appendix C for details).

This article seeks to answer what type of active learning is most appropriate and efficient for physical chemistry and quantum mechanics contexts and how incorporation of active learning elements outside the course lectures impacts learning. Consequently, the focus is on the quantitative analysis of the substantial learning gains obtained with the new course structure in 2017 and the description of the adopted practices and their pedagogical underpinnings. Analysis of the qualitative data on, for example, student attitudes to the introduced changes such as the peer-grading process is left to a subsequent publication.

The rest of the article is organised as follows: Section 2 presents the main challenges in teaching quantum mechanics in chemistry at the university level. The study samples, technical details of the course, and methods are outlined in Section 3. This is followed by an exposition of the employed teaching strategies in Section 4. A discussion of the quantitative results and possible methodological shortcomings is provided in Section 5. The central findings and conclusions are summarised in Section 6.

2 Challenges in teaching of quantum chemistry and spectroscopy

Quantum mechanics and physical chemistry, in general, is distinguished from many other chemistry topics by the large number of abstract concepts whose accurate understanding requires the application of mathematics (Tsaparlis, 2001). Consequently, mathematical ability is an essential factor in predicting students' success in physical chemistry (Derrick and Derrick, 2002; Hahn and Polik, 2004). Indeed, a strong mathematical background forms a necessary but not sufficient condition for success (Derrick and Derrick, 2002; Tsaparlis, 2007). This is because while the transference of skills from mathematics courses into applied subjects is challenging for many students (Sadaghianin, 2005; Thompson *et al.*, 2006; Hadfield and Wieman, 2010; Becker and Towns, 2012) also conceptual understanding and logical thinking skills are crucial for mastering physical chemistry topics. Regardless, for many students in quantum chemistry classes, the challenges of learning new material are exacerbated by gaps in their mathematical knowledge.

In addition to the abstract nature of concepts, Sözbilir (Sözbilir, 2004) found agreement between both students and instructors in that instructor-centred pedagogical approaches, overwhelming amounts of course content, lack of resources, and wilting student motivation all contribute to students' learning difficulties in physical chemistry. Indeed, a significant portion of students enters physical chemistry courses with negative perceptions and low expectations for personal success (Nicoll and Francisco, 2001; Partanen, 2016). Marshman and Singh place the large variations in student motivation, preparedness, and goals along with the paradigm shift necessary for learning quantum mechanics as key factors in explaining student difficulties in physics students' learning of quantum mechanics (Marshman and Singh, 2015). Such negative attitudes are detrimental to student engagement and may act as negative self-fulfilling prophecies even leading students to drop out of the course.

The challenging nature of some of the most central quantum chemical concepts is underlined by the fact that many students both enter and exit undergraduate-level quantum mechanics courses with a myriad of alternative conceptions (Tsaparlis, 2007; Tsaparlis and Papaphotis, 2009). Because quantum mechanical concepts are not regularly encountered in everyday life, these alternative conceptions can arise, for example, from the mixing of various levels of models introduced throughout the students' school career (Taber, 2001) or from the elementary, mostly pictorial, and imprecise treatment of these topics in previous courses (Tsaparlis, 1997). A related way in which difficulties arise is the overgeneralisation of quantum mechanical concepts learned in one context to an inappropriate one (Singh, 2008). Tsaparlis and Papaphotis found that twelfth-grade students possessed several alternative conceptions, claiming that, for example, orbitals represent a definite, well-bounded region in space (Tsaparlis and Papaphotis, 2002). Taber (Taber, 2002a, 2002b) reported that British college students confounded the terms orbital, shell, and orbit; while Harrison and Treagust (Harrison and Treagust, 2000) found confusion in senior high-school students between electron shells and electron clouds. As commented by Taber, it is no wonder that students who struggle with their conceptions already on the level of atomic orbitals face insurmountable learning difficulties when the focus shifts to the treatment of molecular orbitals (Taber, 2002a, 2002b). This is evident in the research findings by Nakiboglu and Zoller who discovered deficiencies in student understanding of hybridisation stemming from alternative conceptions about underlying concepts such as atomic orbitals and the interpretation of s, p, d, and f designations (Zoller, 1990; Nakiboglu, 2003). Recent studies have also found that students have difficulty grasping the basic principles of quantum mechanics, struggling with probability distributions of experiment outcomes (Passante *et al.*, 2015; Marshman and Singh, 2017), and the roles that the time-dependent and -independent Schrödinger equations play in quantum mechanics (Singh and Marshman, 2015).

A striking feature of student alternative conceptions is their ensuing appeal and resistance to change even after instruction. For example, Coll and Treagust together with Coll and Taylor (Coll and Treagust, 2001, 2002; Coll and Taylor, 2002) have demonstrated that even advanced students who were familiar with more abstract models of chemical bonding, preferred simple and realistic ones and employed the advanced models only in the context of tests and examinations. According to Tsaparlis (Tsaparlis, 1997), after passing a compulsory quantum chemistry course, most students failed to provide an exact definition of an atomic orbital. Even after discussions with knowledgeable peers, some failed to align their understanding of quantum mechanical concepts with the scientific view (Tsaparlis and Papaphotis, 2009). Thus, while students may show proficiency in algorithmic processes, such as writing out all Slater determinants arising from a particular electron configuration, they typically do not grasp the underlying principles.

Finally, in contrast to most physics and chemistry courses, in teaching quantum mechanics it is insufficient to specify only the didactic approach and proceed from there to the current scientific understanding of the topic. One also has to consider which philosophical interpretation of quantum mechanics one wishes to follow (Greca and Freire Jr., 2003). That is, while the formalism of quantum mechanics is well established and can be used to make accurate predictions about the behaviour of atoms and molecules, the interpretation of some of its central concepts remain open. For example, the relation of the wave function to the objective reality and the nature of the wave-particle duality change depending on the philosophical framework. While a significant number of scientists can be said to adhere to the Copenhagen interpretation (Bohr, 1935), other incompatible interpretations such as Bohmian mechanics (Bohm, 1952), the many-worlds interpretation (Everett, 1957), the ensemble interpretation (Ballentine, 1970), and the Seoul interpretation (Zhang, 1998) have also gained traction in recent decades. Accordingly, alternative approaches to the teaching of quantum mechanics have been suggested based on these theoretical frameworks (see, for example, (Greca and Freire Jr., 2003; Passon, 2004; Cheong and Song, 2014)). As the choice of the philosophical framework impacts the interpretation of some key experimental results and concepts, it is vital that this issue is given careful thought before the course, and that regardless of which framework is chosen students are made aware where scientific consensus ends, and the instructor's proclivities begin.

3 Methods

3.1 Course background and student assessment

The Structure of molecules and spectroscopy (SMS) course is a five ECTS unit at the University of Helsinki. It is obligatory for chemistry majors in the chemistry Eurobachelor-program, and voluntary for students not majoring in chemistry. The course is intended for the second or third year of studies. In the wake of previous chemistry courses, students should possess a rudimentary understanding of some of the central concepts of the course, such as the wave function, but in general, most of the covered topics are novel. The basic studies in chemistry include a mathematics course (Mathematics for chemists) that aims to prepare students for the physical chemistry courses. In this course, the students practice, for example, integration in spherical coordinates and learn to operate on functions with different operators.

In 2016 and 2017, the course consisted of approximately 35 hours of lectures with each lecture lasting either 90 or 135 minutes. In 2016, the course had six weekly problem bundles, whereas in 2017 this number was dropped to five. The course staff included the author of this article as the teacher-in-charge and two teacher's assistants (TAs), who were either last year bachelor's or master's students at the Department of Chemistry.

One-third of the SMS course grade, or 15 course points, was determined by the weekly exercises with the remaining two-thirds coming from the exam. The full set of 15 course points from the exercises was awarded for getting 90% of the solutions correct. To encourage hard-working students who exceeded this limit up to a maximum of four bonus course points were available based on the exceeding amount.

As the style of exam questions both influences and directs student learning (Entwistle and Entwistle, 1991; Carson and Watson, 2002), I placed significant emphasis on conceptual understanding in 2016 and 2017 with about half of the exam questions tapping into this category. In both years, the students were allowed to bring an A4-size self-written cheat sheet to the exam. The impact of these cheat sheets on learning is debatable, with the majority of current studies indicating no significant effect (Gharib *et al.*, 2012; Hamouda and Shaffer, 2016), and others showing enhanced exam performance (Dickson and Bauer, 2008; de Raadt, 2012). Regardless, cheat sheet exams should be preferred over regular ones because they reduce test anxiety, the results between different kinds open and closed book exams tend to correlate strongly with one another, and because students prefer these exams over closed book ones (Gharib *et al.*, 2012; Hamouda and Shaffer, 2016).

In 2015, the same course unit was organised with a different teacher and 26 lecture hours, following a traditional instructional approach. That year the exercise points obtained from the five problem sets given out during the course did not constitute a part of the grade, but rather depending on the number of problems any given student had solved, a maximum of 4 bonus points could be obtained on top of the 30 course points available from the exam. As in the 2016 and 2017 courses, the students were allowed to bring a cheat sheet with them to the exam.

3.2 Quantitative instruments

The mixed-method approach employed to study the effectiveness of the exercise structure adopted in 2017 made use of three quantitative learning indicators: the percentage of exercise points obtained during the course, the total percentage of points obtained from the course exam, and the results of a conceptual quantum chemistry test (CQCT). The students took the CQCT both at the beginning (pre-CQCT) and at the end of the course (post-CQCT). The CQCT was created by combining and translating into Finnish two carefully validated conceptual quantum mechanics inventories: The Quantum Chemistry Concepts Inventory (QCCI) by Dick-Perez *et al.* (Dick-Perez *et al.*, 2016) and the Quantum Mechanics Conceptual Survey (QMCS) by McKagan *et al.* (McKagan *et al.*, 2010). While other inventories have been developed, (Cataloglu, 2002; Cataloglu and Robinett, 2002; Falk, 2004; Wuttiprom *et al.*, 2009) these two were chosen based on their overlap with central course concepts and a suitable level of difficulty. The final version of the CQCT consisted of 21 questions picked from these two inventories: questions 2–9, 11 and 13 from the QCCI and questions 1–11 from the QMCS. The translation was validated by several faculty staff members and student volunteers who had passed the course in previous years. The questions for QCCI are provided in the supplemental information of the original article by Dick-Perez *et al.* (Dick-Perez *et al.*, 2016), whereas the questions for QMCS are available from the PhysPort-website† upon registration.

Most of the QCCI and QMCS questions were included based on their correspondence with the learning objectives of the course. However, question 12 of the QMCS was excluded because strictly speaking the correct answer this double-slit question depends on which interpretation of quantum mechanics is employed. Question 10 of the QCCI was left out because several example wavefunctions in the course fail to fulfil the properties purportedly required from the wavefunction by this question. For example, a particle in a box wavefunction is not differentiable at the edges of the box where there is an infinite discontinuity of the potential. However, it is not only usable as a wavefunction, it is the energy eigenfunction of the problem in question, implying that more than one of the possible responses were correct for this item.

The pre-CQCT consisted of a subset of 11 questions out of the total 21 items in the post-CQCT. The excluded questions covered topics that had not been discussed in previous courses like the particle in a box, tunneling, and anharmonicity. It was assumed that for these questions the probability of obtaining the correct solution would be almost equal to chance, so given the prevalence of negative preconceptions in students entering the course, the inclusion of these items in the pre-CQCT was deemed counterproductive. The pre-CQCT items were chosen based on consultations with teachers of the preceding courses and students who had recently passed these courses with top grades. The items included in the pre-CQCT were 2, 4, 6, 7, 9, and 11 from the QCCI and 1–5 from the QMCS.

3.3 Study sample and analysis

The official number of enrolled students varied between 50 and 120 depending on the sampling time and the year. To qualify for the quantitative analysis, students had to either have data for two out of the three learning indicators or have accumulated exercise points throughout the course. Of the enrolled students, 56 met these inclusion criteria in 2016 and 60 in 2017. In comparison, the number of students who responded to both pre- and post-CQCT were 52 and 54, respectively. To contrast the results with learning outcomes from a more traditional instructional approach, I also report quantitative data from the 2015 course. As no CQCT was available in 2015, only people who had submitted exercises throughout the course or had done some exercises and participated in the course exam were included in the data that year. This resulted in a sample of 24 students. The quantitative analysis of the course results was conducted using SPSS.‡

3.4 Ethical perspectives

This study was carried out in the spirit of action research (Ralle and Eilks, 2002; Tripp, 2005; Gibbs *et al.*, 2017) where the author of this article served both as the responsible teacher in the SMS course and as the principal researcher. While this type of research is commonplace in educational circles, the dual role of the instructor poses a threat to the validity of the

† PhysPort, https://www.physport.org/assessments/, accessed 7.2.2018.

‡ IBM Corp. Released 2016, IBM SPSS Statistics for Windows, Version 24.0, Armonk, NY: IBM Corp.

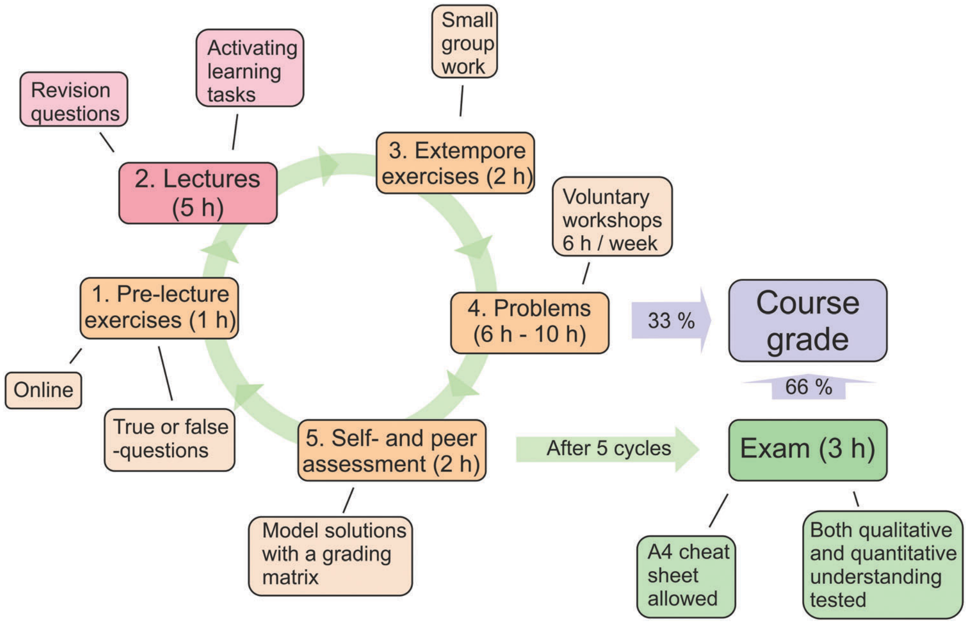

Fig. 1 A summary of the cyclical course structure and assessment practices. Both at the beginning and end of the course, there was one week when only lectures were organised.

data (Nolen and Putten, 2007). The teacher was minimally involved in the grading of course tasks to mitigate these issues. The course exercises were graded either automatically by the Moodle platform, the TAs, or self- and peer-graded by the students based on guidelines set by the instructor. In contrast, one of the five exam questions was personally graded by the instructor due to the limited number of course staff in both years. To avoid any unconscious bias in exam questions, in 2017 two colleagues outside the course as well as the TAs were asked to review and compare both the exam and the retake exam of with the 2016 exams.

This study follows the methodology of my previous research (Partanen, 2016) for which ethical considerations were reviewed by two experts in the field of university pedagogy. It has been also conducted in close collaboration with the university pedagogical assistance staff at the University of Helsinki, providing institutional oversight of the project. Regarding informed consent, during the first lecture students were told that the course results such as grade averages and exercise points would be used in ongoing educational research. Additionally, both the pre- and post-CQCT contained a statement that the results would be employed in pedagogical research. As they constituted but a small fraction of the course tasks, the students had the option of not answering either with negligible effect on course performance.

4 Course tasks, practices, and pedagogical underpinnings

The cyclical course structure and the adopted assessment practices are summarised in Fig. 1. The course lectures, which remained virtually unchanged between 2016 and 2017 are detailed in Appendix A whereas the exercises are described Section 4.1. All course information and materials were available on a Moodle-based online learning environment.§ In 2017, this course area was also used to manage the logistics of the adopted exercise system, such as the random assignment of peer-graders during step 5 in Fig. 1. The textbook in use was Atkin's Physical Chemistry (Atkins and de Paula, 2014), which was complemented by the instructor's notes incorporating material from the educational literature and other popular textbooks (Levine, 2008; Engel and Reid, 2014). In 2016 and 2017, the platform also contained a biweekly updated course diary detailing what topics would be covered in the upcoming lectures and the corresponding pages in the coursebook. It was also updated with full lecture recordings about one day after the relevant lecture. Several discussion forums were available on the platform for students to ask questions from the instructor or TAs, or discuss weekly exercises. The University of Helsinki's Presemo-system¶ was used for anonymous communication between students and the teacher and voting activities during course lectures.

4.1 Facilitating student learning through an effective exercise structure

The new exercise structure adopted in 2017 was based on the 360° feedback theoretical framework by Tee and Pervaiz (Tee and Pervaiz, 2014) with influences from the Assessment cycle-model for peer assessment by Reinholz (Reinholz, 2016). In the 360° feedback model, the goal is to construct a holistic framework of formative assessment to enhance students' learning. The six core elements of this approach are feedback quantity, quality, timing, communication, ability to promote reflection, and social pressure exerted by peers in an environment dominated by

§ Moodle, https://moodle.helsinki.fi/, accessed 7.2.2018.
¶ Presemo, http://presemo.helsinki.fi/, accessed 7.1.2018.

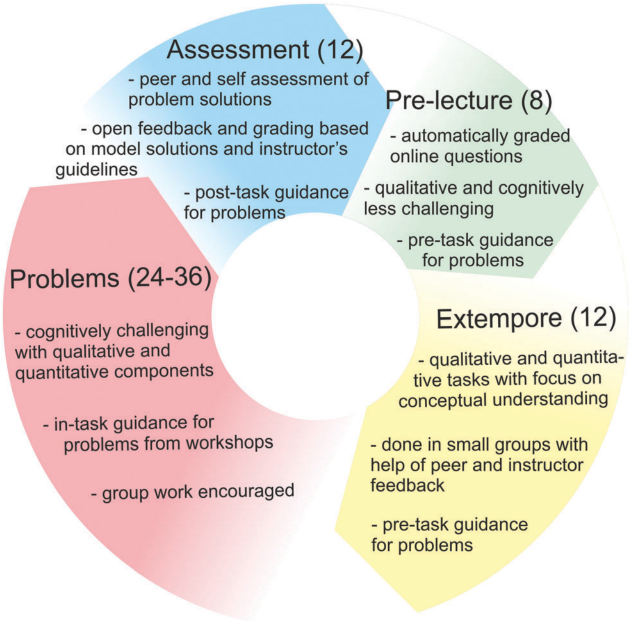

Fig. 2 A pictorial presentation of the four-part exercise structure adopted for the 2017 course. The numbers in parenthesis indicate the number of exercise points available at each part.

positive social interdependence. Incorporation of these elements resulted in the four-stage exercise structure illustrated in Fig. 2. One of the principal objectives was to distribute student effort evenly across the course to ensure that students are continuously guided by feedback which helps them to address alternative conceptions and other issues in performance (Crooks, 1988; Freeman and Lewis, 1998; Gibbs, 1999). The opportunities to give and receive feedback at multiple stages during assessment also help to reorient students' motivation and effort in an appropriate way (Carless, 2002; Gibbs and Simpson, 2004).

Pre-lecture exercises

The first element of the weekly exercise routine was a set of four pre-lecture true or false -questions dispensed biweekly. The questions were simple and focused on the core concepts of the weekly material. The goal was to get students to familiarise themselves with the course material beforehand, thus improving both learning and readiness to discuss related topics in class (Dobson, 2008; Johnson and Kiviniemi, 2009; Moravec *et al.*, 2010). The questions were made available a few days before the relevant lecture, and typically belonged to the lowest two cognitive processing and knowledge categories of Bloom's taxonomy. The students responded to these questions on the course platform, which provided immediate feedback on solutions. The system also yielded valuable information to the lecturer, as he could then use the pre-lecture exercises to assess which parts of the studied topics warranted scrutiny and engage any prevalent issues during the next lecture.

Extempore exercises

The second part of the exercise routine was a two-hour extempore exercise session where the students were given a set of tasks that they then solved in small groups of about 2–4 people with aid from a TA. There were about fifteen students in each session. The solutions were discussed at the end of the class based on the answers offered by the different groups. The sessions aspired to provide timely feedback from both teachers and peers, as appropriate feedback and guidance are necessary components of high-quality practice (Johnson, 2001). In such settings where teamwork is encouraged and where solutions are requested from groups rather than individuals, positive social pressure can propel group members to stay up to beat with others, complete one's share of work, and facilitate learning in others (Partanen, 2016). More generally, collaborative problem-solving enables students to work with others within the same zone of proximal development increasing academic achievement, student attitudes towards the subject, self-esteem, and retention of the material (Springer *et al.*, 1999; Prince, 2004; Lyon and Lagowski, 2008). It also gives the TAs a glimpse into some aspects of the student reasoning that are not apparent from the examination of written solutions, further improving feedback quality.

The extempore exercises were a mix of both qualitative and quantitative questions. On the Revised Bloom's Taxonomy, they were designed to probe and develop understanding mostly at the second through fifth cognitive processing categories and second or third knowledge categories, with many of the questions incorporating elements belonging to the Conceptual knowledge category. Pictorial presentation of wave functions and energy level diagrams and the information that could be gleaned from such images were recurring topics in the exercises. The exercises attempted to make the relationships between different models such as the harmonic and Morse oscillators more salient by asking students to compare and contrast different models. By supporting the formation of multiple representations for a given problem, these tasks also aimed to enhance problem-solving skills in general (Bodner and Domin, 2000; van Heuvelen and Zou, 2001; Madden *et al.*, 2011). For three sample exercises, see Appendix C.1.

Weekly problems

Along with the course lectures, the pre-lecture and extempore exercises were designed to provide pre-task guidance to the most cognitively and mathematically demanding part of the course, *i.e.*, the weekly problems. Three problems and one bonus task were given out each week. The students downloaded their solutions to the course homepage, grading their own and the papers of two peers during the fourth part of the exercise routine. A sample of three problems can be found in Appendix C.2. The problems incorporated tasks that mostly belonged to the second or third knowledge categories and third through fifth categories of the cognitive processing dimension of the Revised Bloom's taxonomy. They were, however, both mathematically and conceptually more demanding than the extempore exercises with a larger fraction belonging to the Procedural knowledge category and the higher-order cognitive process categories 4–6, with some problems even requiring the students to write small essays. Other problems required students to interact with simulations, sketch various functions, derive equations, and look for information online. While the focus was on the quantitative side of quantum mechanics, most problems also included qualitative questions that prompted students to interpret their results in some meaningful way. In an attempt to make the problem-solving process more transparent, complicated calculations were often broken down into different sub-tasks that the students performed in different parts of the problem.

Three approximately two-hour problem-solving workshops were organised with either the lecturer or a TA present to provide in-task guidance and to facilitate positive interdependence through group work. This is important because student engagement is heavily dependent on contextual factors, like relationships between students and teachers (Bryson and Hand, 2007; Price *et al.*, 2011), with the close contact between teacher and student in and out of classes impacting learning more than the actual number of class hours (Gibbs, 2010).

TA preparation

To prepare the TAs for the extempore and workshop classes, I held two weekly meetings among course staff where the solutions, typical student problems, and the pedagogical aspects of teaching were discussed. Rather than just doling out solutions, the TAs were instructed to engage the students in a dialogue, encouraging them to verbalise their solutions and by giving hints and asking questions help individuals and groups overcome challenges on their own. The TAs also received pedagogical training before the course, as preparing the TAs pedagogically for their job has been associated with increased levels of confidence and a better understanding of the difficulties experienced by the students (Romm *et al.*, 2010).

Self- and peer-assessment

The fourth part of the exercise structure consisted of students' self- and peer-assessment of the problem solutions based on a grading matrix and model solutions. The students were also required to give written feedback on the other students' solutions, for example by justifying why they had deducted points in a given exercise or about the general features of the other student's solutions like readability. Clear instructions on how to do the self- and peer-assessment were provided in the course platform to improve the quality of this post-task guidance. As mentioned, each student was tasked with assessing their own paper and those of two randomly assigned peers. To enhance the validity of the assessment, the actual grade was calculated as an average of the three numbers. In cases where the minimum and maximum grades differed substantially, the course instructor provided a definitive grading. As recommended by Ballantyne (Ballantyne *et al.*, 2002), the students remained anonymous to each other throughout the assessment process and received credit based on the number and quality of their assessments. This kind of peer assessment against model solutions with a specific grading matrix is a highly valid way of assessing exercises when compared with instructor markings (Falchikov and Boud, 1989; Stefani, 1994; Dochy *et al.*, 1999; Falchikov and Goldfinch, 2000).

In addition to being resource effective, peer- and self-assessment has been associated with a number of benefits for the student. With a careful choice of the assessed tasks, increased exposure can improve student learning and problem-solving skills (Topping, 1998; Dochy *et al.*, 1999; Pereira *et al.*, 2016), while the assessment process develops critical thinking, assessment, and reflection skills (Topping, 1998; Dochy *et al.*, 1999; Pereira *et al.*, 2016). Placing students in an active role within the assessment process also increases perceptions of fairness and effectiveness (Flores *et al.*, 2015), as well as motivation and involvement in the learning process (Stefani, 1994; Pereira *et al.*, 2016; López-Pastor and Sicilia-Camacho, 2017). This is especially true if the assessment process enables an ongoing dialogue between the student and the instructor, so I tried to provide the students with both anonymous and non-anonymous ways of contacting course staff regarding the grading matrices or model solutions, but only a handful ever did.

Exercise systems in 2016 and 2015

In 2016, the exercises consisted only of pre-lecture assignments similar to the 2017 ones and a set of harder problems, whereas in 2015 the pre-lecture assignments were also absent. In both years, student solutions were marked by the TA. In 2016, she also provided more extensive feedback on one pre-agreed exercise per week. In both 2016 and 2015, an open workshop was organised at least six hours per week. The TAs received a few hours of pedagogical training before the course in 2016, and a weekly meeting was organised to go through the solutions of the weekly problems and the grading guidelines. Compared to 2017, these meetings were more focused on the subject matter and not on the pedagogical aspects of teaching.

5 Results and discussion

Table 1 summarises the learning outcomes of the SMS course from the three instruments of this study. Based on preliminary correlation analysis, the exam result was seen to correlate with both exercise and CQCT scores in both years with correlation

Table 1 Summary of the student learning outcomes measured by the three different learning indicators in 2016 and 2017 together with exam results and exercise point data from 2015. The symbol N is the sample size

	Exam result (%)	Exercise points (%)	Pre-CQCT score (%)	Post-CQCT score (%)	Post-CQCT (full) score (%)
2017					
Average	66.3	66.9	63.1	72.1	67.7
Std. dev.	(19.6)	(17.7)	(20.4)	(17.4)	(15.3)
N	57	60	54	54	57
2016					
Average	48.4	60.5	61.0	64.5	56.2
Std. dev.	(18.1)	(17.7)	(19.8)	(17.3)	(15.1)
N	54	56	52	52	55
2015					
Average	48.4	59.7	—	—	—
Std. dev.	(11.6)	(16.1)	—	—	—
N	19	24	—	—	—

coefficients above 0.4. This is relieving, as it can be taken as a signal that the exam measures both qualitative and quantitative understanding of course topics. Thus, it can be considered as a more general measure of student learning than the CQCT. In both years, also the pre-CQCT correlated significantly with the exam result, meaning that students who entered the course with a solid background from previous courses generally performed well in this one. As described in Section 3, the year 2015 has been included in the analysis as a comparison point for results obtained with a more classical type of instruction.

5.1 Learning outcomes

Comparing years 2017 and 2016 in Table 1, one sees a substantial improvement in both the exam results and the student ability to answer the post-CQCT, while the number of exercise points increases only slightly. The statistical significance of these changes was investigated using an independent samples t-test with a 95% confidence interval using a two-tailed distribution. For the exam result, the difference was statistically significant ($t(109) = 5.005$, $p = 0.000$) with an estimated Cohen's d effect size of $d = 0.95$. For the full post-CQCT, the difference is also statistically significant ($t(110) = 4.000$, $p = 0.000$) with an effect size of $d = 0.76$. According to the interpretation guidelines suggested by Sawilowsky (Sawilowsky, 2009) both of these d values correspond to a large effect. In contrast, the difference in the percentage of exercise points is not statistically significant ($t(114) = 1.950$, $p = 0.054$). Since the total time required for the exercises was approximately equal, and the course lectures were similar in both years, this suggests that the new exercise structure maximized the efficiency of learning by requiring students to exert more cognitive effort in the same time period. However, this conclusion is based on the assumption that the student populations are initially comparable in both years. To investigate this claim, I looked at the pre-CQCT scores of the students in both 2016 and 2017 using an independent samples t-test, which found no significant difference between the populations' initial knowledge or ability regarding course topics ($t(104) = 0.542$, $p = 0.589$). The null hypothesis was further consolidated by looking at the exam results of those students in the 2016 and 2015 prerequisite Mathematics for chemists course, where no statistical difference was observed between the two different classes ($t(122) = 0.515$, $p = 0.607$).

The improvements observed in Table 1 are congruent with the results obtained by Deslauriers and Wieman (Deslauriers and Wieman, 2011) who compared the effectiveness and impact on long-term retention of interactive instruction strategies to traditional approaches at a modern physics course at the University of British Columbia using QMCS. The interactively taught class scored 19% better than the one under traditional instruction while both showed little deterioration in their QMCS score when tested 6 or 18 months after the end of the course. More generally, the results align with the emerging consensus within the natural sciences that active learning approaches improve learning (Prince, 2004; Ruiz-Primo *et al.*, 2011; Freeman *et al.*, 2014). As predicted by Freeman *et al.* (Freeman *et al.*, 2014), the effect sizes observed in this kind of second-generation action research, where active learning elements are present also outside the lectures, are larger than the $d = 0.47$ value reported in their meta-analysis.

The complexity of the final course structure makes it challenging to pinpoint exactly how much different features impacted student learning. However, based on the sizeable observed effect size between 2016 and 2017 and the partial overlap of the respective exercise systems, it seems likely that the incorporation of the extempore exercises and the peer- and self-assessment of the problems were mostly responsible for the improvements in learning. This conclusion is supported by the gains observed in conceptual understanding through the introduction of an extempore exercise system in my previous study (Partanen, 2016).

In contrast to previous findings (Prince, 2004; Ruiz-Primo *et al.*, 2011; Freeman *et al.*, 2014), Table 1 shows no substantial difference between the 2015 and 2016 results. Thus, the 2015 and 2016 results are considered here as a baseline for comparing the new instructional approaches of the 2017 course. It should be noted, however, that this surprising lack of difference might stem from changes in course goals and exam questions between 2015 and 2016 that camouflage any learning gains obtained from the incorporation of active learning elements to lectures. These methodological issues are further explored in Section 5.5.

5.2 Within course comparison of the pre- and post-CQCT results

One measure of the learning that occurred within the SMS course each year can be obtained by looking at student responses to the subset of CQCT questions included in the pre-CQCT, both at the beginning and at the end of the course. The average results are given in the third and fourth columns of Table 1. Curiously, while some improvement is seen in student answers for these questions in 2016, a paired-samples t-test gives no statistical significance between the results ($t(51) = 1.509$, $p = 0.137$). As seen from the table, in 2017 this difference is larger and reaches statistical significance ($t(53) = 3.925$, $p = 0.000$) with $d = 0.47$ corresponding to a medium-sized effect. These changes translate to normalised gain values of $G = 0.24$ in 2017 and $G = 0.09$ in 2016. However, it should be borne in mind that this pre- and post-CQCT comparison only captures a part of the learning because the CQCT measures only conceptual understanding and because almost half of the questions of the full CQCT were not present in the pre-CQCT. This issue is discussed more thoroughly in Section 5.5.

Because there were substantial differences in learning in the pre- and post-CQCT results between the two years, it is worthwhile to take a closer look at these CQCT outcomes. In Tables 2 and 3 the ranked four percentile groups of both the pre- and post-CQCTs are tabulated based on the binning of the pre-CQCT. The categories of the pre- and post-CQCTs are represented by the rows and columns of the matrices, respectively. Improvement in the results can be seen as the accumulation of students in the third and fourth columns of the tables, *i.e.*, in the above average and highest-categories, whereas the No-change ones are in the diagonal.

It is clear that in Table 2 more students rank into the highest percentile group relative to the starting results than in Table 3.

Table 2 Tabulation of ranked percentile groups of the pre- and post-CQCTs of the 2017 course showing the number of students and the mean point differences ($\bar{x}$) in each cell (N = 54)

Post-CQCT	Lowest n = 7 ($\bar{x}$)	Below average n = 16 ($\bar{x}$)	Above average n = 11 ($\bar{x}$)	Highest n = 20 ($\bar{x}$)
Pre-CQCT				
Lowest n = 12	5 (1.8)	4 (2.5)	1 (5.0)	2 (5.5)
Below average n = 21	2 (−1.5)	9 (0.0)	8 (1.4)	2 (3.5)
Above average n = 8	0 —	1 (−1.0)	1 (0.0)	6 (1.7)
Highest n = 13	0 —	2 (−2.5)	1 (−1.0)	10 (0.0)

Table 3 Tabulation of ranked percentile groups of the pre- and post-CQCTs of the 2016 course showing the number of students and the mean point differences ($\bar{x}$) in each cell (N = 52)

Post-CQCT	Lowest n = 5 ($\bar{x}$)	Below average n = 15 ($\bar{x}$)	Above average n = 18 ($\bar{x}$)	Highest n = 14 ($\bar{x}$)
Pre-CQCT				
Lowest n = 10	2 (1.0)	5 (1.8)	3 (3.7)	0 —
Below average n = 15	2 (−1.5)	7 (0.0)	4 (1.8)	2 (3.0)
Above average n = 13	1 (−4.0)	2 (−2.5)	5 (0.2)	5 (1.4)
Highest n = 14	0 —	1 (−5.0)	6 (−1.2)	7 (0.1)

In both years, more students are located in the upper-right corner of the matrix than in the lower-left, but in 2017 the percentage of these students is larger. The more substantial difference, however, is the percentage of students going down in their learning category, located in the lower-left corner of the matrix. In 2016, the number of these students was 12 while in 2017 it was only 6. This could mean that in 2016 a significant portion of students got good results from the pre-CQCT through guessing or, more alarmingly, that the instruction itself contained elements that confused and misled students, leaving them less informed in some regards than they were before taking the course. I tried to answer this question by looking at the exercise and exam points of the students who scored lower on their post-CQCT than pre-CQCT. A number of students who scored two or more points lower on their post-test showed decreasing trends in exercise points towards the end of the course and either had not participated in the exam or had scored very low. It seems that these students had become disengaged with the course and thus may not have exerted as much effort when responding to the post-CQCT as they did for the pre-CQCT, hoping to just quickly obtain the exercise points awarded from completing the posttest. In other cases, the students were actively engaged throughout, obtaining average results for both the final exam and the post-CQCT, but their pre-CQCT score was remarkably high. Here the decreased test result can likely be explained by fortuitous guessing during the pretest. There were also cases where some detrimental effects from instruction could not be ruled out. For example, in both years some high-performing students who had obtained almost full points from the pre-CQCT scored one or two points lower on the post-CQCT. Typically, these students had corrected the pre-CQCT answers they had gotten wrong, but had accumulated new alternative conceptions in the post-CQCT. Alternatively, it could also be that they had guessed these items correctly in during the pre-CQCT and failed to do so for the post-CQCT.

Another noteworthy feature of Tables 2 and 3 is that despite the apparent successes of the new course structure, the majority of students in both years are located on the diagonal, implying that there was no substantial change between their pre- and post-CQCT score. Indeed, if one looks at the mean point differences in the diagonal categories they are all less than equal to one. This underlines the deep-rooted nature of many of the alternative conceptions demonstrated in earlier studies (Tsaparlis, 1997; Coll and Treagust, 2001, 2002; Coll and Taylor, 2002; Tsaparlis and Papaphotis, 2009), but it also highlights the relative difficulty of the original QCMS and QCCI instruments. Again, it should be noted that the pre-CQCT consisted only of a subset of the CQCT questions that the students could answer based on previous courses, which likely decreased the difference between the post- and the pre-CQCT scores.

5.3 Question specific differences in the CQCT

To shed further light into the question specific differences in the CQCTs, the stacked bar plots for the pre-CQCT responses of both 2016 and 2017 are shown in Fig. 3. The 2016 post-CQCT results are shown in Fig. 4, whereas the 2017 results are given in Fig. 5. Comparing the pre-CQCT percentages of the two years in Fig. 3, one sees larger than 10% differences in four questions. Importantly, half the time better results are obtained in 2016, underlining that no significant differences exist in initial student competence between the two years.

Comparison of the pre- and post-CQCT results within 2016 and 2017

Looking at Fig. 3 and 4, one sees that in only four of the questions (Q2, Q8, Q15, and Q16) the result of the post-CQCT is more than 10% better than the result of the pre-CQCT, signifying that the improvement in conceptual understanding was somewhat moderate in 2016. In contrast, for Q10 there is a 20% decrease in the number of correct answers after instruction. This question was a simple true or false-statement, where the claim was that free electrons follow sinusoidal trajectories in the absence of external forces. As seen from the figure, in both 2016 pre- and post-tests, the majority of students seemed to confuse the functional form of the wave function with the electron's motion. In hindsight, this increase in the prevalence of alternative conceptions is not surprising, given that many of the wave functions in the course have a sinusoidal form and that there was almost no discussion about the trajectories of particles, only about the probability density. Likely this negligence prevented students from forming a clear image of the relationship

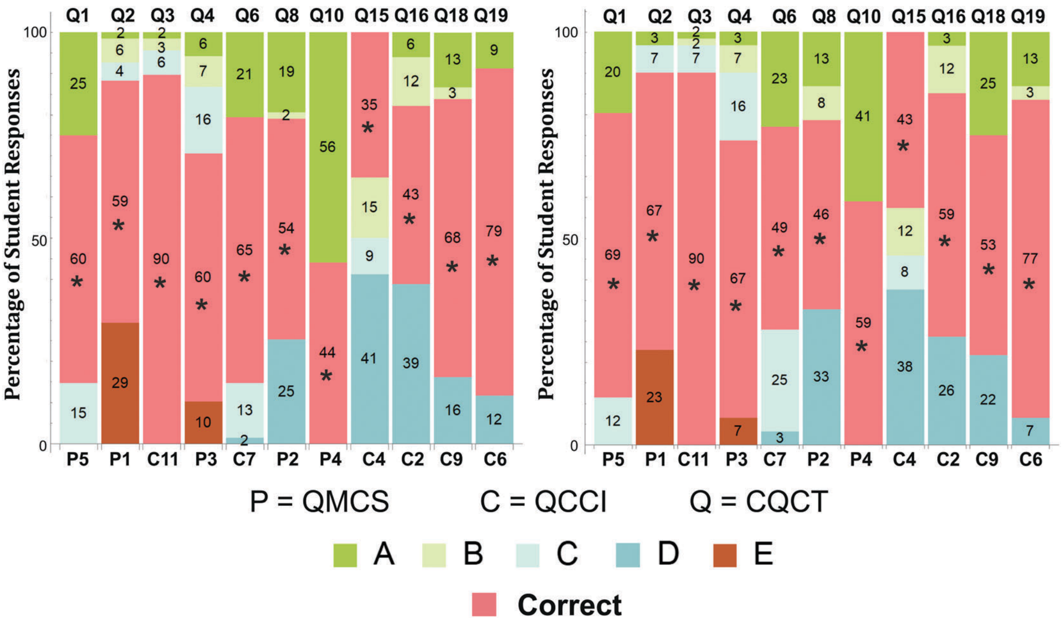

Fig. 3 Student responses as a percentage choosing each answer for the pre-CQCT in 2016 (left) and 2017 (right). The correct answer is marked by the red color and an asterisk in each case. The symbols Q1, Q2,. . . refer to the numbering employed in the CQCT whereas P1, P2,. . . and C1, C2,. . . refer to the corresponding questions in QMCS and QCCI, respectively. Due to rounding errors, the percentages in each stack might not add up to exactly 100.

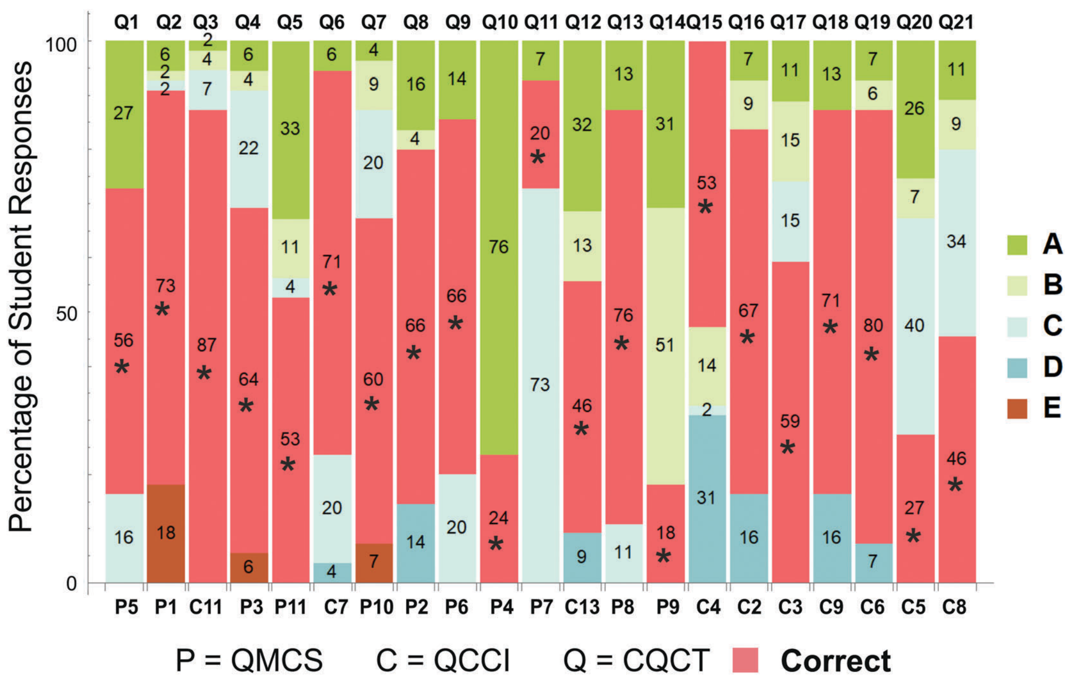

Fig. 4 Student responses as a percentage choosing each answer for the post-CQCT in 2016. The correct answer is marked by the red color and an asterisk in each case. The symbols Q1, Q2,. . . refer to the numbering employed in the CQCT whereas P1, P2,. . . and C1, C2,. . . refer to the corresponding questions in QMCS and QCCI, respectively.

between the particle trajectories galore in classical mechanics and the probabilistic view suggested by the Born interpretation. It also partially explains the larger number of students going down in their knowledge category in Table 3 compared to Table 2. In contrast to 2016, the greater conceptual understanding between the pre- and the post-CQCT observed in 2017 is evident in Fig. 3 and 5, where for all questions included in the pre-CQCT, the results of the post-test are either practically equal or, in most cases, better.

Comparison of the post-CQCT results between 2016 and 2017

Compared to the post-CQCT percentages of 2016, the 2017 results show more than 30% increases in the number of correct

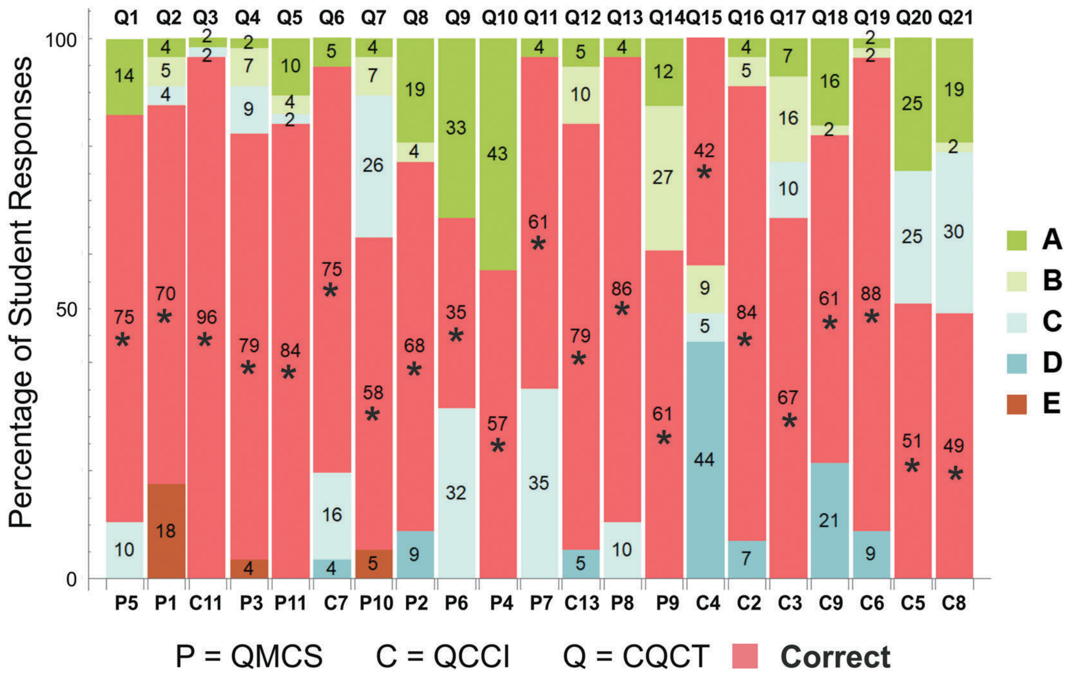

Fig. 5 Student responses as a percentage choosing each answer for the post-CQCT in 2017. The correct answer is marked by the red color and an asterisk in each case. The symbols Q1, Q2,... refer to the numbering employed in the CQCT whereas P1, P2,... and C1, C2,... refer to the corresponding questions in QMCS and QCCI, respectively.

answers for five questions (Q5, Q10, Q11, Q12, and Q14) highlighting the significant gains in student conceptual understanding between the two years. The opposite trend is seen only for Q9 which had to do with consequent measurements of two non-commuting variables (energy and position) for the particle in a box, and the collapse of a wave function upon measurement. This is curious, as since this question was identified as challenging after the 2016 course, the relevant parts of the course lectures were restructured by incorporating additional student activities and a more careful consideration of how the wave function changes when complementary variables are measured in sequence. In light of the 30% decrease in the number of correct answers, these changes seem to only have confused students further, with each of the response options being about equally popular in the post-CQCT of 2017.

Regarding the problematic question Q10 of 2016, 57% of students answered correctly in 2017. The improvement is likely because this alternative conception was challenged multiple times in 2017, for example, by explicitly requesting students to describe what they could say about the motion of an electron as part of the extempore exercises after they had been asked to draw the wave function of a particle. Still, about the same number of students that started the course with the alternative conception that the electrons move along sinusoidal trajectories left it with their preconception intact.

Questions Q11 and Q14 are examples where changes to the activating tasks in the lectures and topic-focused extempore exercises resulted in significant gains in student understanding. Of these two, Q11 had to do with what happens to the energy of a particle when it encounters a potential wall and tunnels through it. As expected based on classical mechanics, a significant portion of students erroneously claimed that it loses energy. In 2016, this number was 73% whereas in 2017 only about 35% of students had this alternative conception. This issue was mentioned in the course material in 2016, but during the 2017 lectures, it was investigated through a voting and discussion activity focused on the potential energy surface for the inversion of ammonia, asking how this change in geometry affected the energy of the molecule. This discussion also tied in to the factors affecting the probability of tunneling like barrier width and height.

Akin to Q11, Q14 showed a dismal original result. This question asked what can be said about the allowed energies of a particle in a finite box when its energy is higher than the walls of the box. In 2016, 51% percent of students falsely claimed that the energy would be quantised whereas in 2017 the number of correct answers increased to 61%. To effect this change, the lecture material pertaining to the particle in a box model was revised and a separate module was devoted to the qualitative features of the finite box. For example, this module included a voting activity where the students had to consider what happened to the de Broglie wavelength of the particle in different regions of the 1-dimensional box when the initial energy of the particle was greater than the depth of the potential energy well. In addition to this increased focus during the lectures, the second extempore exercise also included a task where the students sketched the energy level diagram and the wavefunctions of the different energy levels for a particle in a finite box. Unfortunately, despite the successes in Q11 and Q14, still a substantial number of students answer incorrectly, indicating that further development of material is mandated for these challenging topics.

The impact of the new exercise structure is most apparent in Q12, which asked students to connect approximate wavenumber of a photon to a corresponding type of transition, *i.e.* electronic, rotational, or vibrational. In contrast to 2016, when a new transition type came under focus in 2017 either the extempore exercises or problems always asked the students to compare it to previously learned transitions and to see how the two relate to each other through an energy level diagram. This principle is illustrated in Example problem III of Appendix C.2, where after calculating the ionisation energy of hydrogen, the student is asked to compare it to the dissociation energy of H_2 to get an idea of the energy scales involved in these familiar processes. As a result of this change in focus, the number of students who answered correctly increased from 46% to 79% between the two years.

Like Q9, Q15 was an example where intervention failed to have a significant impact on student learning or backfired altogether. It asked how the energy of a system changes when a chemical bond is formed. In 2016, around one-third of students claimed that when a bond forms, energy may be absorbed or emitted depending on the atoms involved in the process. Despite having encountered numerous molecular orbital diagrams demonstrating that the total energy of the molecule is lower than that of the constituent atoms, and having been asked to spell out this rationale numerous times during the course, the percentage of students having the same alternative conception increased to 44% in 2017. A possible explanation to this is that many students were still missing the link between stability or total energy of a molecule and the concept of bond formation.

Finally, questions Q5 and Q20 give an idea of the impact that extempore exercises had on student learning. Question Q5 showed a plot of a one-dimensional wavefunction and asked the student to judge the probabilities of finding the particle in different regions based on this graph. As illustrated in Example exercises I and III in Appendix C.1, the connection between the wavefunction and particle location was a recurring theme in the extempore exercises, as it was intricately tied to virtually all of the major topics of the course. In this light, it is hardly surprising that the number of correct responses increased 31% when moving from 2016 to 2017.

Question Q20 had to do with the connection between classical and quantum mechanical harmonic oscillator through the correspondence principle. As illustrated in Example exercise II of Appendix C.1, as part of the extempore exercises about the quantum harmonic oscillator, the students were for example asked to compare the behaviour of this model to its classical counterpart. As seen from Fig. 4 and 5, this resulted in a 24% increase in the number of correct answers between 2016 and 2017. However, as with questions Q11 and Q14, despite clear gains in learning only about half of the students were able to answer correctly at the end of the course with several distractors accumulating a substantial percentage of student responses.

5.4 Comparison of CQCT responses with original samples

It is also illustrative to compare the pre- and post-CQCT results of this study to the percentages of correct responses reported in the original publications of Dick-Perez *et al.* (Dick-Perez *et al.*, 2016) for the QCCI and McKagan *et al.* (McKagan *et al.*, 2010) for the QMCS. Of these two, only Dick-Perez reported data before instruction. Based on this data, it appears that the students participating in the SMS course may possess a slightly better prior grasp of quantum mechanics than the undergraduate North American chemistry students that were surveyed in the original QCCI sample. In five of the six QCCI questions incorporated in the pre-CQCT, the Finnish students performed more than 10% better than the American ones in 2017, whereas in 2016 this was true for three of the questions. Regarding the post-CQCT, in 2017 in 9 out of the included 10 QCCI-items, the percentage of correct answers for the Finnish students was higher than the North American ones, and in four of these cases this difference was larger than 10%. In 2016, this was true for only five of the ten items with four showing more than 10% increase.

McKagan *et al.* tested their conceptual inventory on modern physics students (McKagan *et al.*, 2010) who likely possess superior mathematics skills compared to the chemistry students of this study. The different backgrounds of chemistry and physics students make the comparison of results somewhat challenging. Regardless, of the 11 QMCS questions included in the CQCT, in 2017, the Finnish chemistry students obtained better results in six, with a difference of more than 10% in the number of correct answers in four of these. For 2016, these numbers were four and three. These results indicate that the course structure employed in 2017 yields improved learning results not only compared to the 2016 course, but also to the more general standard represented by the relevant populations of these studies. Finally, factor analysis was employed to investigate whether the post-CQCT results tended to load onto factors aligned with the initial instruments, but no such tendency was observed.

5.5 Threats to validity

The use of multiple instruments for measuring learning outcomes in 2016 and 2017 increases the persuasiveness of the results even though the relatively large variations in student gains in all three years raise some concerns about the reliability of the findings. On the other hand, as the number of students and sample demographics were similar, and as the results of the pre-CQCT and previous years' Mathematics for chemists class showed no statistically significant difference, it is unlikely that there has been much variation between the 2016 and 2017 student samples in starting knowledge or average academic ability. Furthermore, the observed differences are statistically significant even at the $p = 0.01$ level and show large effect sizes.

At face value, the surprising lack of improvement in the exam results between 2015 and 2016 might indicate that the move from the teacher-centred to the student-centred lecture structure did not significantly improve student learning, implying that the learning gains are heavily dependent on the precise nature of the adopted approach. However, this change also lead to a revision of the course objectives and a corresponding increase in the difficulty of the exam as more questions tapping into the higher cognitive processing categories of Bloom's taxonomy were included. The low number of participants in the 2015 course, and the change in the course instructor in addition to the pedagogical approach raise concerns about the comparability of these results with the 2016 and 2017 ones. Furthermore, it should be noted that in 2016 there was a partial overlap between one of the obligatory laboratory courses and the SMS lectures, which may have negatively impacted engagement for a portion of students. However, as the course lectures were available on video at the course platform in both 2016 and 2017, it is unlikely that this factor had a substantial adversary effect on the 2016 results. Despite the issues associated with the 2015 class, it was the most comparable traditionally taught class of the past quantum mechanics courses at the Faculty of Science, because the course underwent major changes prior to 2015 due to a curricular overhaul. Even in the light of these issues, it seems safe to conclude that the 2017 structure was superior to both the 2016 and the traditional 2015 ones.

In light of the larger learning gains observed in 2017 compared to 2016, it seems surprising that the improvement between the 2017 pre- and post-CQCT results remains relatively small. One explaining factor might be that, as described in Section 3.2, only questions to which the students could know the answer before the course were included in the pre-CQCT. This lead to an exclusion of approximately half of the CQCT questions. This is not the standard practice when administering conceptual tests such as the Force concept inventory (FCI) (Hestenes *et al.*, 1992), where the full form is presented both at the beginning and end of the course and the normalised gains scores are consequently higher (Coletta and Phillips, 2005). I chose not to follow this practice because it might reinforce the low expectations of success many students possess at the beginning of physical chemistry courses. Indeed, McKagan *et al.* (McKagan *et al.*, 2010) found that administering the QMCS as a pretest to modern physics students was both demoralizing and pointless, as the results were practically equal to chance.

Finally, the dual role of the author as both course instructor and researcher is always a potential threat to the reliability of action research. As described in Section 3.4 many precautions were taken to overcome possible ethical concerns by, for example, distancing the instructor from the actual grading process as much as possible and using external sources to validate the comparability of the exams on different years.

6 Summary and conclusions

This study answers the call by Freeman *et al.* (Freeman *et al.*, 2014) for second-generation research on active learning. Its purpose is to elucidate what type of methodology is most appropriate and efficient for a given context and student population, and how incorporation of specific active learning elements outside the course lectures impacts learning. This was done in the context of physical chemistry and specifically in quantum mechanics and spectroscopy by proposing a student-centred course and exercise structure based on current pedagogical research. The effectiveness of the new structure was measured using three quantitative learning indicators: exercise points obtained by the students, exam results, and the results of the pre- and post-CQCTs.

In addition to the activating learning tasks such as discussion questions and voting activities in the course lectures in 2016, in 2017 the course included a four-stage student-centred exercise structure summarised in Fig. 2. This inclusion substantially improved learning compared to 2016 and the traditionally taught 2015 class. As predicted by Freeman *et al.* (Freeman *et al.*, 2014), the effect sizes observed here are larger than the average ones reported in their meta-analysis for the impact of active learning on student performance in science, engineering and mathematics. This highlights the benefits of incorporating elements from active learning not just to lectures, but also to other course tasks. Together with my previous action research study on thermodynamics (Partanen, 2016), this study demonstrates the benefits of a multifaceted, student-centred approach to both exercises and lecturing in courses where quantitative and qualitative understanding is essential for a deep-rooted knowledge of the subject. Specifically to quantum mechanics, the results align with those obtained by Deslauriers and Wieman (Deslauriers and Wieman, 2011) in that interactively taught classes show improved learning when compared to traditional ones. In the broader picture, this study provides yet another data point to the mounting evidence (Prince, 2004; Ruiz-Primo *et al.*, 2011; Freeman *et al.*, 2014) supporting the effectiveness of active learning approaches in natural sciences.

Conflicts of interest

There are no conflicts of interest to declare.

Appendices

A Description of adopted lecture practices

The 2016 and 2017 courses utilised a broken lecture structure. These structures are characterised by regular interruptions of the lecture by activating learning tasks (Knight and Wood, 2005), which improve student attention (Bunce *et al.*, 2010) and facilitate a broad range of learning outcomes (Prince, 2004). The activating learning tasks also increase satisfaction and the perceived effectiveness of the lectures (Smith, 2006; Miller *et al.*, 2013). The largest difference between the broken structures adopted in my earlier study (Partanen, 2016) and the ones here was the greater amount of time devoted to the activating learning tasks, especially in 2017.

The material was divided into a series of modules of approximately 25 to 60 minutes in duration with about 10 minute breaks between modules. In contrast, traditionally a two or three-hour lecture in Finland consists, respectively, of two or three 45 minute sections with 15 minute breaks between each section. The topics of the modules are given in Appendix B. Each module delved into one of the central themes of the course.

The activating tasks within the modules included quantitative exercises, discussion questions, multiple choice voting, and drawing assignments. A principal goal was to get students to work on their answers in small groups, engaging them in collaborative learning (Prince, 2004). The tasks varied annually, as they were modified based on alternative conceptions encountered in the pre-CQCT or previous lectures. For example, in 2017, at the start of the module covering the hydrogen atom, students were asked to draw a depiction of hydrogen as accurately as they could. These sketches were gathered, and a discussion question was created based on student submissions at the beginning of the next lecture. The tasks were often based on items employed in the literature to probe alternative conceptions (see Harrison and Treagust for the above example (Harrison and Treagust, 2000)). The activating learning tasks allow the instructor to engage alternative conceptions promptly, and adjust the material and tasks of the future lectures accordingly. As demonstrated by the Peer Instruction literature (Mazur, 1997; Crouch and Mazur, 2001; Meltzer and Mannivannan, 2002; Lasry *et al.*, 2008; Smith *et al.*, 2009; Turpen and Finkelstein, 2009; Smith *et al.*, 2011), this results in a marked improvement in student conceptual understanding and quantitative problem solving ability.

At the beginning of each module, students were also tasked to find answers to between three to five review questions. The questions were visible throughout the module, were closely aligned with its central learning goals, and served as a starting point for review at the end of each lecture. They forced the students to immediately apply the material under study, facilitating open discussion and the identification of alternative conceptions and difficult concepts in the course material (Knight and Wood, 2005). As such, the review questions constituted another channel of direct feedback on student learning.

Concerning the philosophical interpretation of quantum mechanics, the suspensive perspective suggested by Cheong and Song (Cheong and Song, 2014) was adopted for the 2016 and 2017 lectures. In this approach, a distinction is made between the prediction rules and the reality claims rather than between the formalism and interpretation of quantum mechanics. The prediction rules contain not only the formalism but also noncontroversial parts of this interpretation, such as the utilisation of wave function collapse as a calculation tool. In short, these rules consist of a set of equations and calculation procedures for the prediction of phenomena. The reality claims include the normative propositions relating theory to reality. In the suspensive perspective, the prediction rules are taught as accepted scientific knowledge, while the plethora of competing reality claims are only superficially discussed with no preference to one over the others.

A different lecturer was responsible for the course in 2015. The lectures followed a more traditional structure in which the lecturer first presented the students with the information and then occasionally showed how to apply the formulae in practice by going through some example calculations. Due to the rapid pace endemic to the traditional lecturing approaches, this year many additional topics were covered.

B Course content

The names of the modules covered in each lecture of the course in the autumn of 2017 have been translated from Finnish into English below. In 2015 and 2016, the same topics were covered, but as described in Appendix A, between each year there were substantial differences in how these topics were presented. As is common when comparing student and instructor centred courses, many less relevant issues were also discussed in the 2015 classes. In addition to the items listed below, the lectures started with an introductory module where the course structure, tasks, and learning goals were discussed with the students.

- Fundamentals of quantum mechanics

1. Farewell to classical mechanics
2. Basic properties of the wave function
3. Operators in quantum mechanics
4. Heisenberg uncertainty principle
5. Superpositions and measurement in quantum mechanics
6. Review: Postulates of quantum mechanics

- Simple quantum mechanical systems with one particle

1. Particle in an infinite box
2. Particle in a finite box and applications
3. Particle and a potential wall—the curious case of tunneling

- Two-particle systems

1. Harmonic oscillator
2. Interlude: Introduction to spectroscopy
3. Vibrational spectroscopy of diatomic molecules
4. Review: Classical rotational motion
5. Basics of quantum mechanical rotation
6. Rotational spectroscopy of diatomic molecules
7. Hydrogenic atoms

- Many-particle systems

1. Many-electron atoms
2. Spin
3. Computational chemistry of atoms
4. Molecular orbital theory for the hydrogen molecule
5. Molecular orbital theory for diatomic molecules
6. Molecular orbital theory for polyatomic molecules
7. Rotational, vibrational and electronic spectroscopy for polyatomic molecules

C Examples of course tasks

The extempore exercises and course problems shown below were translated from Finnish into English. Each task has been classified based on Bloom's taxonomy, as indicated by the red letters inside parentheses. The letter k refers to the knowledge dimension and the letter c to the cognitive process dimension in the taxonomy. In the knowledge dimension, the different subcategories have been numbered as 1. factual, 2. conceptual, 3. procedural, 4. metacognitive. In the cognitive process category, the subcategories have been numbered as 1. remember, 2. understand, 3. apply, 4 analyze, 5. evaluate, and 6. create. So, for example, (k2c4) would correspond to Knowledge category: conceptual knowledge, Cognitive process category: analyze. A similar classification of chemistry Matriculation Examination problems has been performed by Tikkanen and Aksela

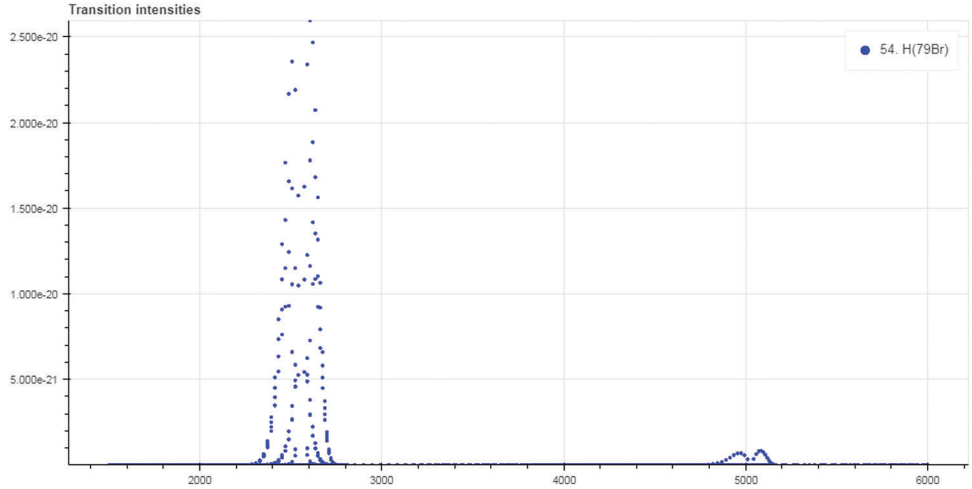

Fig. 6 Spectrum for the $H^{79}Br$ in cm^{-1}.

(Tikkanen and Aksela, 2012), and more detailed information about the classification process can be found therein. As the classifications reported here are based on the judgements of a single individual, they should be considered approximate.

C.1 Sample extempore exercises

Example exercise I. Intro: In this extempore exercise we are going to delve into the intricacies of the Particle in a box model! The electrons in conjugated π electron systems present in dye molecules like 1,6-diphenyl-1,3,5-hexatriene can be described surprisingly accurately as particles in a one-dimensional box. When 1,6-diphenyl-1,3,5-hexatriene absorbs electromagnetic radiation at 375 nm an electron is excited from the highest occupied state $n = 3$ to $n = 4$.

1. Draw the wave functions for the electron in the ground $n = 3$ and excited $n = 4$ states. Based on your drawings, in which state does the electron possess higher kinetic energy and why? (k3c4)

2. Let us now look closer at the electron in the excited state $n = 4$.

(a) Calculate the expectation value of the electron position $\langle x \rangle$. (k3c3)

(b) Calculate the most probable location of finding the electron. (k3c3)

(c) Mark the quantities you obtained in (a) and (b) into the graph of exercise 1. How can you deduce the expectation value and the most probable location directly from your graph in this case? Determine these properties graphically for $n = 3$. (k3c4)

Example exercise II. Intro: In this extempore session, we will explore the $H^{79}Br$ spectrum taken from the Hitran database,‖ where on the horizontal axis we have the wavenumber in cm^{-1} (Fig. 6).

2. Let's start interpreting the figure by first treating HBr as a harmonic oscillator:

(a) The peak on the left corresponds to the transition $\nu = 0 \leftarrow 1$. Use this information to find the force constant of HBr and compare it to the 1860 N m^{-1} force constant of CO. (k3c3)

(b) Based on the results you obtained in (a), sketch the potential energy graph and the allowed energy levels for the HBr molecule. Draw the transition used in (a) to your picture. What kind of spectrum you would expect for HBr based on your figure and the harmonic oscillator selection rules? Interpret the provided Hitran spectrum in light of your own energy level diagram. (k2c5)

(c) Sketch the probability density $\|\psi\|^2$ for the two lowest energy levels into your energy level diagram from (b). Using these, list four differences between a quantum mechanical harmonic oscillator and a classical harmonic oscillator where two spherical objects have been coupled by a spring. (k3c4)

Example exercise III. Intro: In this extempore exercise we will look at the simplest molecular systems: H_2^+ and H_2. For H_2^+ a good approximation for the two lowest molecular orbitals is obtained from a LCAO-MO of the form

$$\psi_{\pm} = N(\psi_{1s_A} \pm \psi_{1s_B})$$

1. Let's start by looking at H_2^+:

(a) Sketch the ψ_{1s} orbitals shown in the superposition into a single graph so that both nuclei A and B lie on the horizontal axis, equal distance away from the origin. In the ground state the internuclear distance is $R_e = 1.05$ Å. (k3c2)

(b) Sketch two additional graphs containing the $\psi_{\pm}$ molecular orbitals. Based on your drawings, which molecular orbitals is bonding and which is antibonding? Why? (k3c4)

(c) How will the molecular orbitals look if you increase the internuclear distance tenfold? Compare the probability densities of the antibonding and bonding orbitals in this case. (k3c4)

(d) Explain what is meant by the variational principle. How would you expect the energy of the ground state molecular orbital change if 2s and 2p orbitals were added to the above LCAO-MO? (k3c3)

C.2 Sample course problems

Example problem I. This problem is particularly instructive in terms of upcoming course topics! A frequently employed approximation for the Hamiltonian of the vibrational motion of diatomic molecules like HCl is of the form encountered in

‖ Hitran online (Gordon *et al.*, 2017), http://hitran.org/, accessed 14.1.2018.

problem 4: $\hat{H} = \frac{\hat{p}^2}{2\mu} + \frac{1}{2}k_f x^2$, where x is the deviation from the equilibrium bond length, μ is the reduced mass, and k_f is the force constant of the bond. Let us try to guess the ground state wave function for HCl by looking at a Gaussian function $\psi(x) = Ne^{-ax^2}$, where $a > 0$.

(a) Write the Schrödinger equation for the system. (k3c3)

(b) Calculate the normalization constant for the guessed function. Is $\psi(x)$ a possible wave function? (*I.e.*, does it fulfil the properties required from a wave function?) (k3c4)

(c) By operating with $\hat{H}$, find out the value for a at which $\psi(x)$ fulfils the Schrödinger equation. What energy value corresponds to this wave function? (k3c3)

(d) BONUS: Following the examples in the extempore exercises, calculate $\langle \hat{x} \rangle$ and the most probable deviation from the equilibrium bond distance for the ground state of HCl. Give a physical interpretation of the numbers you obtain as a result of these calculations. Sketch the probability density and use that to explain why the values are equal or non-equal. (k3c5) (*Hint:* If you run into nasty-looking integrals, you should search for integration formulas online!)

(*This problem has been modified from original problems in Atkins' Physical Chemistry* (Atkins and de Paula, 2014)).

Example problem II. In this exercise we try to explain the resonance stabilization of the conjugated π-electron systems encountered in extempore 2 based on the particle in a box model. You should start the problem by playing around with the QuVis-simulation for the particle in a box, which can be found at https://www.st-andrews.ac.uk/physics/quvis/simulations_hhtml5/sims/Particles-infwell/particles-infwell.html

Let's model the π-electrons in butadiene either as localised to two separate π-bonds of length L or as delocalised over the whole molecule as in the resonance model. Let's assume that the electrons do not interact in either case (Fig. 7).

(a) What are the quantum numbers corresponding to the highest occupied energy level (Highest occupied molecular orbital, HOMO) and the lowest unoccupied energy level (Lowest unoccupied molecular orbital, LUMO) in the **resonance model**? Sketch the relevant molecular energy level diagram. Remember to place the π-electrons into your diagram! (k3c3)

(b) Evaluate the plausibility of the resonance model by comparing the energies of the localised and delocalised models. (k3c4)

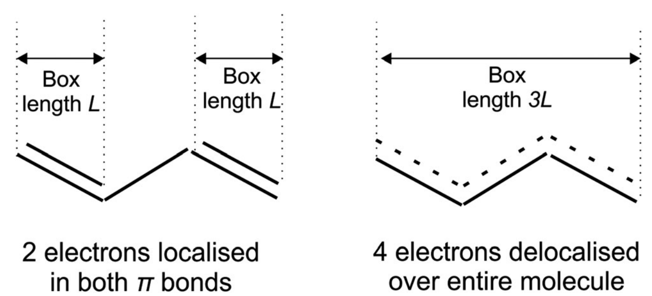

Fig. 7 The localised and delocalised models of bonding in butadiene.

(c) In the spectrum of butadiene, the maximum wavelength at which absorption occurs is 217 nm. Draw the measured transition into the energy level diagram sketched in (a) and calculate the average π-bond length for butadiene in the resonance model. Compare your result to the literature values 1.336 Å (C–C bond) and 1.454 Å (C=C bond). Explain any possible differences. (k3c5)

(*Hint:* In reality the π-electron network extends beyond the atomic nuclei. When calculating the average bond length, account for this by using the number of bonds +1 in the denominator!)

(*This problem has been modified from original problems in Engel's Physical Chemistry* (Engel and Reid, 2014) *and deSouza and Iyengar* (deSouza and Iyengar, 2013)).

Example problem III. Positronium consists of an electron and a positron orbiting around their common center of mass. The principal features of its spectrum are expected to be hydrogen-like so that the biggest differences arise from the mass differences.

(a) Calculate the reduced mass for positronium and compare it with the reduced mass of hydrogen. Describe the motions of the positively and negatively charged particles in the system on the basis of your calculations. (k3c4)

(b) Calculate and compare the ionization energies of positronium and hydrogen. For further comparison, calculate also the dissociation energy for H_2. Based on extempore 3 the relevant Morse oscillator parameters are ω_e = 4405.3 cm^{-1} and $x_e\omega_e$ = 125.325 cm^{-1}. (k3c4)

(c) Hydrogen is the most abundant element in stars. Based on problem 2 and your solution to (b), why are no absorption or emission lines of hydrogen present in the spectrum of stars with effective temperatures higher than 25 000 K? (k2c4)

(*This problem has been modified from original problems in Atkins' Physical Chemistry* (Atkins and de Paula, 2014)).

Acknowledgements

I thank Dr Liisa Myyry, and Dr Anni Rytkönen for invaluable suggestions. I also would like to thank Prof. Lauri Halonen and Prof. Kari Laasonen for support and allowing me to conduct this research as part of my teaching and research work in the University of Helsinki and Aalto University. Finally, I thank Prof. Sari Lindblom and Prof. Maija Aksela for suggestions and comments. I also thank my course assistants Timo Pekkanen, Matias Jääkeläinen, Iiris Tuoriniemi, and Miisamari Jeskanen for a plethora of useful suggestions and improvements for the course material and exercises.

References

Adams W. K., Reid S., LeMaster R., McKagan S. B., Perkins K. K., Dubson M. and Wieman C. E., (2008a), A study of educational simulations part i – engagement and learning, *J. Interact. Learn. Res.*, **19**, 397–419.

Adams W. K., Reid S., LeMaster R., McKagan S. B., Perkins K. K., Dubson M. and Wieman C. E., (2008b), A study of educational simulations part ii – interface design, *J. Interact. Learn. Res.*, **19**, 551–577.

Anderson L. W. and Krathwohl D. R., (ed.), (2001), *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, New York: Longman.

Atkins P. and de Paula J., (2014), *Atkins' Physical Chemistry*, 10 edn, Oxford, New York: Oxford University Press.

Ballantyne R., Hughes K. and Mylonas A., (2002), Developing procedures for implementing peer assessment in large classes using an action research process, *Assess. Eval. High. Educ.*, **27**, 427–441.

Ballentine L. E., (1970), The statistical interpretation of quantum mechanics, *Rev. Mod. Phys.*, **42**, 358–381.

Becker N. and Towns M., (2012), Students' understanding of mathematical expressions in physical chemistry contexts: An analysis using Sherin's symbolic forms, *Chem. Educ. Res. Pract.*, **13**, 209–220.

Bloom B. S., (1956), *Taxonomy of educational objectives*, Michigan: Edwards Bros, Ann Arbor.

Bodner G. M. and Domin D. S., (2000), Mental models: the role of representations in problem solving in chemistry, *Univ. Chem. Educ.*, **4**, 24–30.

Bohm D., (1952), A suggested interpretation of the quantum theory in terms of "hidden" variables. ii, *Phys. Rev.*, **85**, 180–193.

Bohr N., (1935), Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, **48**, 696–702.

Bryson C. and Hand L., (2007), The role of engagement in inspiring teaching and learning, *Innov. Educ. Teach. Int.*, **44**, 349–362.

Bunce D. M., Flens E. A. and Neiles K. Y., (2010), How long can students pay attention in class? A study of student attention decline using clickers, *J. Chem. Educ.*, **87**, 1438–1443.

Carless D. R., (2002), The 'mini-viva' as a tool to enhance assessment for learning, *Assess. Eval. High. Educ.*, **27**, 353–363.

Carson E. M. and Watson J. R., (2002), Undergraduate students' understandings of entropy and gibbs free energy, *Univ. Chem. Educ.*, **6**, 4–12.

Cataloglu E., (2002), *Development and validation of an achievement test in introductory quantum mechanics: The quantum visualization instrument QMVI*, PhD thesis, The Pennsylvania State University.

Cataloglu E. and Robinett R. W., (2002), Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career, *Am. J. Phys.*, **70**, 238–251.

Cheong Y. W. and Song J., (2014), Different levels of the meaning of wave-particle duality and a suspensive perspective on the interpretation of quantum theory, *Sci. Educ.*, **23**, 1011–1030.

Coletta V. P. and Phillips J. A., (2005), Interpreting FCI scores: normalized gain, preinstruction scores, and scientific reasoning ability, *Am. J. Phys.*, **73**, 1172–1182.

Coll R. K. and Taylor N., (2002), Mental models in chemistry: senior chemistry students' mental models of chemical bonding, *Chem. Educ. Res. Pract.*, **3**, 175–184.

Coll R. K. and Treagust D. F., (2001), Learners' mental models of chemical bonding, *Res. Sci. Educ.*, **31**, 357–382.

Coll R. K. and Treagust D. F., (2002), Exploring Tertiary Students' Understanding of Covalent Bonding, *Res. Sci. Technol. Educ.*, **20**, 241–267.

Crooks T. J., (1988), The impact of classroom evaluation practices on students, *Rev. Educ. Res.*, **58**, 438–481.

Crouch C. H. and Mazur E., (2001), Peer instruction: ten years of experience and results, *Am. J. Phys.*, **69**, 970–977.

de Raadt M., (2012), Student created cheat-sheets in examinations: Impact on student utcomes, Proceedings of the Fourteenth Australasian Computing Education Conference ACE2012. CRPIT, vol. 123, pp. 71–76.

Derrick M. E. and Derrick F. W., (2002), Predictors of success in physical chemistry, *J. Chem. Educ.*, **79**, 1013–1016.

Deslauriers L. and Wieman C., (2011), Learning and retention of quantum concepts with different teaching methods, *Phys. Rev. ST Phys. Educ. Res.*, **7**, 010101.

deSouza R. T. and Iyengar S. S., (2013), Using quantum mechanics to facilitate the introduction of a broad range of chemical concepts to first-year undergraduate students, *J. Chem. Educ.*, **90**, 717–725.

Dick-Perez M., Luxford C. J., Windus T. L. and Holme T., (2016), A quantum chemistry concept inventory for physical chemistry classes, *J. Chem. Educ.*, **93**, 605–612.

Dickson K. L. and Bauer J. J., (2008), Do students learn course material during crib sheet construction? *Teach. Psychol.*, **35**, 117–120.

Dobson J. L., (2008), The use of formative online quizzes to enhance class preparation and scores on summative exams, *Adv. Physiol. Educ.*, **32**, 297–302.

Dochy F., Segers M. and Sluijsmans D., (1999), The use of self-, peer and co-assessment in higher education: a review, *Stud. High. Educ.*, **24**, 331–350.

Engel T. and Reid P., (2014), *Physical Chemistry*, 3rd edn, Harlow, United Kingdom: Pearson Education Limited.

Entwistle N. J. and Entwistle A., (1991), Contrasting forms of understanding for degree examinations: the student experience and its implications, *High. Educ.*, **22**, 205–227.

Everett H., (1957), "Relative state" formulation of quantum mechanics, *Rev. Mod. Phys.*, **29**, 454–462.

Falchikov N. and Boud D., (1989), Student self-assessment in higher education: a meta-analysis, *Rev. Educ. Res.*, **59**, 395–430.

Falchikov N. and Goldfinch J., (2000), Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks, *Rev. Educ. Res.*, **70**, 287–322.

Falk J., (2004), *Developing a quantum mechanics concept inventory*, PhD thesis, Uppsala University.

Finkelstein N. D., Adams W. K., Keller C. J., Kohl P. B., Perkins K. K., Podolefsky N. S., Reid S. and LeMaster R., (2005), When learning about the real world is better done virtually: a study of substituting computer simulations for laboratory equipment, *Phys. Rev. ST Phys. Educ. Res.*, **1**, 010103.

Flores M. A., Veiga Simão A. M., Barros A. and Pereira D., (2015), Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education, *Stud. High. Educ.*, **40**, 1523–1534.

Freeman R. and Lewis R., (1998), *Planning and implementing assessment*, London: Kogan Page.

Freeman S., Eddy S. L., McDonough M., Smith M. K., Okoroafor N., Jordt H. and Wenderoth M. P., (2014), Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 8410–8415.

Galvez E. J., Holbrow C. H., Pysher M. J., Martin J. W., Courtemanche N., Heilig L. and Spencer J., (2005), Interference with correlated photons: Five quantum mechanics experiments for undergraduates, *Am. J. Phys.*, **73**, 127–140.

Gharib A., Phillips W. and Mathew N., (2012), Cheat sheet or open-book? A comparison of the effects of exam types on performance, retention, and anxiety, *Psychol. Res.*, **2**, 469–478.

Gibbs G., (1999), *Assessment matters in higher education*, Buckinham, United Kingdom: Open University Press, ch. Using assessment strategically to change the way students learn, pp. 41–53.

Gibbs G., (2010), *Dimensions of quality*, York, United Kingdom: The Higher Education Academy.

Gibbs G. and Simpson C., (2004), Conditions under which assessment supports students' learning, *Learn. Teach. High. Educ.*, **1**, 3–31.

Gibbs P., Cartney P., Wilkinson K., Parkinson J., Cunningham S., James-Reynolds C., Zoubir T., Brown V., Barter P., Sumner P., MacDonald A., Dayananda A. and Pitt A., (2017), Literature review on the use of action research in higher education, *Educ. Action Res.*, **25**, 3–22.

Gordon I. E., Rothman L. S., Hill C., Kochanov R. V., Tan Y., Bernath P. F., Birk M., Boudon V., Campargue A., Chance K. V., Drouin B. J., Flaud J.-M., Gamache R. R., Hodges J. T., Jacquemart D., Perevalov V. I., Perrin A., Shine K. P., Smith M.-A. H., Tennyson J., Toon G. C., Tran H., Tyuterev V. G., Barbe A., Császár A.G., Devi V. M., Furtenbacher T., Harrison J. J., Hartmann J.-M., Jolly A., Johnson T. J., Karman T., Kleiner I., Kyuberis A. A., Loos J., Lyulin O. M., Massie S. T., Mikhailenko S. N., Moazzen-Ahmadi N., Müller H.S.P., Naumenko O. V., Nikitin A. V., Polyansky O. L., Rey M., Rotger M., Sharpe S. W., Sung K., Starikova E., Tashkun S. A., Vander Auwera J., Wagner G., Wilzewski J., Wcislo P., Yu S. and Zak E. J., (2017), The hitran2016 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, **203**, 3–69, HITRAN2016 Special Issue.

Greca I. M. and Freire Jr. O., (2003), Does an emphasis on the concept of quantum states enhance students' understanding of quantum mechanics? *Sci. Educ.*, **12**, 541–557.

Hadfield L. C. and Wieman C. E., (2010), Student interpretations of equations related to the first law of thermodynamics, *J. Chem. Educ.*, **87**, 750–755.

Hahn K. E. and Polik W. F., (2004), Factors influencing success in physical chemistry, *J. Chem. Educ.*, **81**, 567–572.

Hamouda S. and Shaffer C. A., (2016), Crib sheets and exam performance in a data structures course, *Comput. Sci. Educ.*, **26**, 1–26.

Harrison A. G. and Treagust D. F., (2000), Learning about atoms, molecules, and chemical bonds: A case study of multiple-model use in grade 11 chemistry, *Sci. Educ.*, **84**, 352–381.

Hestenes D., Wells M. and Swackhamer G., (1992), Force concept inventory, *Phys. Teach.*, **30**, 141–158.

Johnson M., (2001), Facilitating high quality student practice in introductory physics, *Phys. Educ. Res. Am. J. Phys. Suppl.*, **69**, 1–11.

Johnson B. C. and Kiviniemi M. T., (2009), The effect of online chapter quizzes on exam performance in an undergraduate social psychology course, *Teach. Psychol.*, **36**, 33–37.

Johnson B. R., Onwuegbuzie A. J. and Turner L. A., (2007), Toward a definition of mixed methods research, *J. Mix. Methods Res.*, **1**, 112–133.

Kalkanis G., Hadzidaki P. and Stavrou D., (2003), An instructional model for a radical conceptual change towards quantum mechanics concepts, *Sci. Educ.*, **87**, 257–280.

Knight J. K. and Wood W. B., (2005), Teaching more by lecturing less, *Cell Biol. Educ.*, **4**, 298–310.

Kohnle A., Douglass M., Edwards T. J., Gillies A. D., Hooley C. A. and Sinclair B. D., (2010), Developing and evaluating animations for teaching quantum mechanics concepts, *Eur. J. Phys.*, **31**, 1441–1455.

Lasry N., Mazur E. and Watkins J., (2008), Peer Instruction: from harvard to the two-year college, *Am. J. Phys.*, **76**, 1066–1069.

Levine I. N., (2008), *Physical Chemistry*, 6th edn, New York: McGraw-Hill Higher Education.

López-Pastor V. and Sicilia-Camacho A., (2017), Formative and shared assessment in higher education. Lessons learned and challenges for the future, *Assess. Eval. High. Educ.*, **42**, 77–97.

Lyon D. C. and Lagowski J. J., (2008), Effectiveness of facilitating small-group learning in large lecture classes, *J. Chem. Educ.*, **85**, 1571–1576.

Madden S. P., Jones L. L. and Rahm J., (2011), The role of multiple representations in the understanding of ideal gas problems, *Chem. Educ. Res. Pract.*, **12**, 283–293.

Marshman E. and Singh C., (2015), Framework for understanding the patterns of student difficulties in quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.*, **11**, 020119.

Marshman E. and Singh C., (2017), Investigating and improving student understanding of the expectation values of observables in quantum mechanics, *Eur. J. Phys.*, **38**, 045701.

Mazur E., (1997), *Peer Instruction: A User's Manual*, New Jersey: Prentice Hall.

McKagan S. B., Perkins K. K., Dubson M., Malley C., Reid S., LeMaster R. and Wieman C. E., (2008), Developing and researching phet simulations for teaching quantum mechanics, *Am. J. Phys.*, **76**, 406–417.

McKagan S. B., Perkins K. K. and Wieman C. E., (2010), Design and validation of the Quantum Mechanics Conceptual Survey, *Phys. Rev. Phys. Educ. Res.*, **6**, 020121.

Meltzer D. and Mannivannan K., (2002), Transforming the lecture-hall environment: the fully interactive physics lecture, *Am. J. Phys.*, **70**, 639–654.

Miller C. J., McNear J. and Metz M. J., (2013), A comparison of traditional and engaging lecture methods in a large, professional-level course, *Adv. Physiol. Educ.*, **37**, 347–355.

Moravec M., Williams A., Aguilar-Roca N. and O'Dowd D. K., (2010), Learn before lecture: a strategy that improves learning outcomes in a large introductory biology class, *CBE Life Sci. Educ.*, **9**, 473–481.

Müller R. and Wiesner H., (2002), Teaching quantum mechanics on an introductory level, *Am. J. Phys.*, **70**, 200–209.

Nakiboglu C., (2003), Instructional misconceptions of turkish prospective chemistry teachers about atomic orbitals and hybridization, *Chem. Educ. Res. Pract.*, **4**, 171–188.

Nicoll G. and Francisco J. S., (2001), An investigation of the factors influencing student performance in physical chemistry, *J. Chem. Educ.*, **78**, 99–102.

Nolen A. L. and Putten J. V., (2007), Action research in education: addressing gaps in ethical principles and practices, *Educ. Res.*, **36**, 401–407.

Partanen L., (2016), Student oriented approaches in the teaching of thermodynamics at universities – developing an effective course structure, *Chem. Educ. Res. Pract.*, **17**, 766–787.

Passante G., Emigh P. J. and Shaffer P. S., (2015), Examining student ideas about energy measurements on quantum states across undergraduate and graduate levels, *Phys. Rev. ST Phys. Educ. Res.*, **11**, 020111.

Passon O., (2004), How to teach quantum mechanics, *Eur. J. Phys.*, **25**, 765–769.

Pereira D., Flores M. A. and Niklasson L., (2016), Assessment revisited: a review of research in assessment and evaluation in higher education, *Assess. Eval. High. Educ.*, **41**, 1008–1032.

Podolefsky N. S., Perkins K. K. and Adams W. K., (2010), Factors promoting engaged exploration with computer simulations, *Phys. Rev. ST Phys. Educ. Res.*, **6**, 020117.

Price M., Handley K. and Millar J., (2011), Feedback: focusing attention on engagement, *Stud. High. Educ.*, **36**, 879–896.

Prince M., (2004), Does active learning work? A review of the research, *J. Eng. Educ.*, **93**, 223–231.

Ralle B. and Eilks I., (ed.), (2002), *Research in chemical education – What does this mean?*, Aachen, Germany: Shaker, ch. Participatory action research within chemical education, pp. 87–98.

Reinholz D., (2016), The assessment cycle: a model for learning through peer assessment, *Assess. Eval. High. Educ.*, **41**, 301–315.

Romm I., Gordon-Messer S. and Kosinski-Collins M., (2010), Educating young educators: a pedagogical internship for undergraduate teaching assistants, *CBE Life Sci. Educ.*, **9**, 80–86.

Ruiz-Primo M. A., Briggs D., Iverson H., Talbot R. and Shepard L. A., (2011), Impact of undergraduate science course innovations on learning, *Science*, **331**, 1269–1270.

Sadaghianin H., (2005), *Conceptual and mathematical barriers to students learning quantum mechanics*, PhD thesis, Columbus, Ohio: The Ohio State University, Electronic Thesis or Dissertation. Retrieved from https://etd.ohiolink.edu/18.12.2017.

Sánches Gómez P. J. and Martín F., (2003), Quantum *vs.* "classical" chemistry in university chemistry education: A case study of the role of history in thinking the curriculum, *Chem. Educ. Res. Pract.*, **4**, 131–148.

Sawilowsky S. S., (2009), New effective size rules of thumb, *J. Mod. Appl. Stat. Methods*, **8**, 597–599.

Sözbilir M., (2004), What makes physical chemistry difficult? Perceptions of Turkish chemistry undergraduates and lecturers, *J. Chem. Educ.*, **81**, 573–578.

Singh C., (2008), Student understanding of quantum mechanics at the beginning of graduate instruction, *Am. J. Phys.*, **76**, 277–287.

Singh C. and Marshman E., (2015), Review of student difficulties in upper-level quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.*, **11**, 020117.

Slunt K. M. and Giancarlo L. C., (2004), Student-centered learning: a comparison of two different methods of instruction, *J. Chem. Educ.*, **81**, 985–988.

Smith D. K., (2006), Use of the mid-lecture break in chemistry teaching: a survey and some suggestions, *J. Chem. Educ.*, **83**, 1621–1624.

Smith M. K., Wood W. B., Adams W. K., Wieman C., Knight J. K., Guild N. and Su T. T., (2009), Why peer discussion improves student performance on in-class concept questions, *Science*, **323**, 122–124.

Smith M. K., Wood W. B., Krauter K. and Knight J. K., (2011), Combining peer discussion with instructor explanation increases student learning from in-class conceptual questions, *CBE Life Sci. Educ.*, **10**, 55–63.

Springer L., Stanne M. E. and Donovan S. S., (1999), Effects of small-group learning on undergraduates in science, mathematics engineering and technology: a meta-analysis, *Rev. Educ. Res.*, **69**, 21–51.

Stefani L. A. J., (1994), Peer, self and tutor assessment: relative reliabilities, *Stud. High. Educ.*, **19**, 69–75.

Taber K. S., (2001), Building the structural concepts of chemistry: Some considerations from educational research, *Chem. Educ. Res. Pract.*, **2**, 123–158.

Taber K. S., (2002a), Conceptualizing quanta: Illuminating the ground state of student understanding of atomic orbitals, *Chem. Educ. Res. Pract.*, **3**, 145–158.

Taber K. S., (2002b), Compounding quanta: probing the frontiers of student understanding of molecular orbitals, *Chem. Educ. Res. Pract.*, **3**, 159–173.

Tee D. D. and Pervaiz K. A., (2014), 360 degree feedback: an integrative framework for learning and assessment, *Teach. High. Educ.*, **19**, 579–591.

Thompson J. R., Bucy B. R. and Mountcastle D. B., (2006), Assessing student understanding of partial derivatives in thermodynamics, Proceedings of the 2005 Physics Education Research Conference of the American Institute of Physics.

Tikkanen G. and Aksela M., (2012), Analysis of Finnish chemistry matriculation examination questions according to cognitive complexity, *Nord. Stud. Sci. Educ.*, **8**, 258–268.

Topping K., (1998), Peer assessment between students in colleges and universities, *Rev. Educ. Res.*, **68**, 249–276.

Tripp D., (2005), Action research: a methodological introduction, *Educ. Pesqui.*, **31**, 443–466.

Tsaparlis G., (1997), Atomic orbitals, molecular orbitals and related concepts: conceptual difficulties among chemistry students, *Res. Sci. Educ.*, **27**, 271–287.

Tsaparlis G., (2001), Towards a meaningful introduction to the Schrödinger equation through historical and heuristic approaches, *Chem. Educ. Res. Pract.*, **2**, 203–213.

Tsaparlis G., (2007), *Advances in teaching physical chemistry, ACS symposium series*, Washington, DC: American Chemical Society, ch. Teaching and learning physical chemistry: a review of educational research, vol. 973, pp. 75–112.

Tsaparlis G. and Papaphotis G., (2002), Quantum-chemical concepts: Are they suitable for secondary students? *Chem. Educ. Res. Pract.*, pp. 129–144.

Tsaparlis G. and Papaphotis G., (2009), High-school students' conceptual difficulties and attempts at conceptual change: the case of basic quantum chemical concepts, *Int. J. Sci. Educ.*, **31**, 895–930.

Turpen C. and Finkelstein N. D., (2009), Not all interactive engagement is the same: variations in physics professors' implementation of Peer Instruction, *Phys. Rev. ST Phys. Educ. Res.*, **5**, 020101.

van Heuvelen A. and Zou X., (2001), Multiple representations of work-energy processes, *Am. J. Phys.*, **69**, 184–194.

Weimer M., (2002), *Learner-centered teaching: five key changes to practice*, San Fransisco, CA: Jossey-Bass.

Wright G. B., (2011), Student-centered learning in higher education, *Int. J. Teach. Learn. High. Educ.*, **23**, 92–97.

Wuttiprom S., Sharma M. D., Johnston I. D., Chitaree R. and Soankwan C., (2009), Development and use of a conceptual survey in introductory quantum physics, *Int. J. Sci. Educ.*, **31**, 631–654.

Zhang H. I., (1998), Epistemic subject and epistemological structure of science, *Korean J. Philos. Sci.*, **1**, 1–33.

Zoller U., (1990), Students' misunderstandings and misconceptions in college freshman chemistry (general and organic). *J. Res. Sci. Teach.*, **27**, 1053–1065.